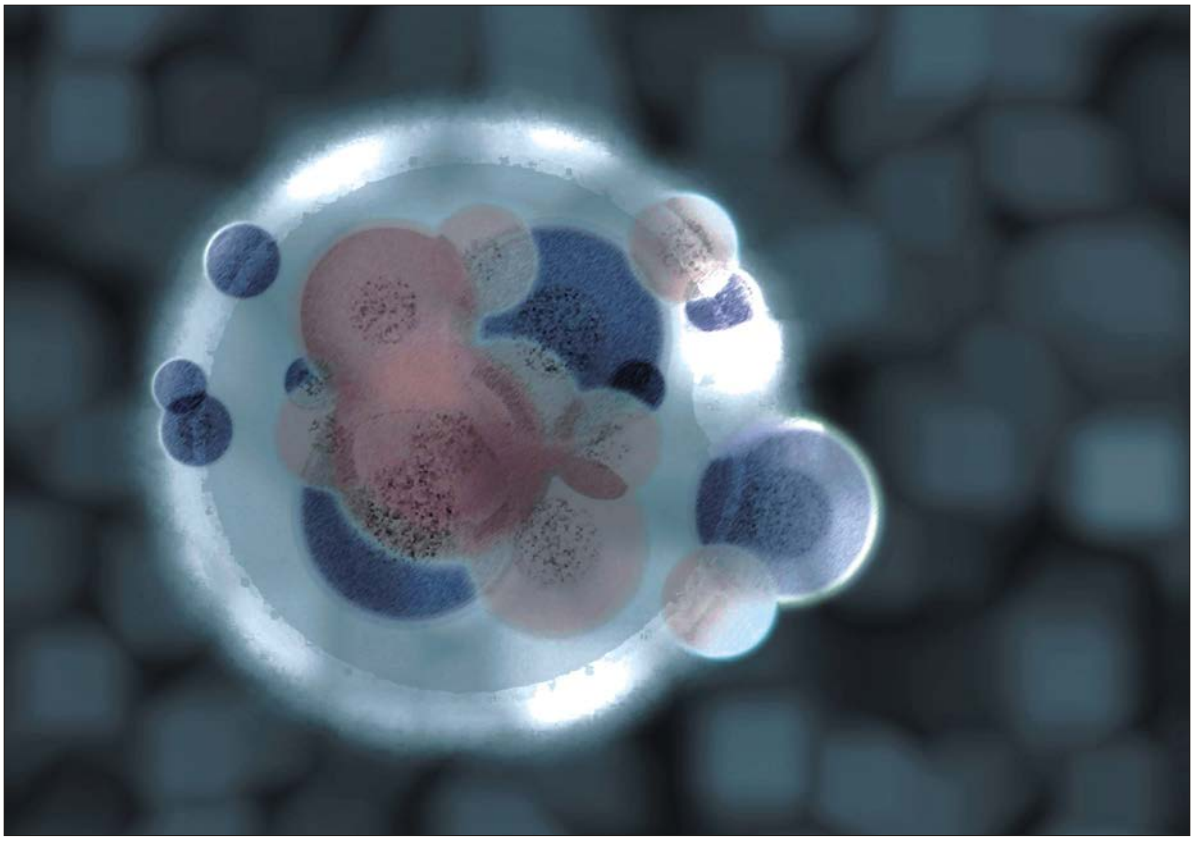

Journal of Biological Sciences

ISSN 1727-3048

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Biological Sciences 8 (8): 1343-1347, 2008
ISSN 1727-3048

Cloning and Characterization of Three B Hordein Genes From *Hordeum vulgare* ssp. *agriocrithon*

[1]Z.E. Pu, [2]Y.C. Hou, [1]X.X. Xu, [2]Z.H. Yan, [2]Y.M. Wei, [2]X.J. Lan and [2]Y.L. Zheng
[1]Agronomy College, Sichuan Agricultural University, Yaan, 625014, China
[2]Triticeae Research Institute, Sichuan Agricultural University, Dujiangyan, 611830, China

Abstract: Three novel B hordein genes (designated as *Ha1*, *Ha2* and *Ha3*) were isolated from the genomic DNA of *Hordeum vulgare* ssp. *agriocrithon* accession ZYM0834 by PCR amplification. The coding regions of *Ha1*, *Ha2* and *Ha3* were 900, 945 and 975 bp, respectively. The deduced amino acid sequences were 297, 312 and 322 amino acid residues each with a signal peptide, a central repetitive region rich in proline and glutamine and C-terminal non-repetitive domains. Seven (for *Ha1* and *Ha2*) and eight (for *Ha3*) cysteine residues were found in the C-terminal domain. In order to investigate the three novel B hordein genes belong to B1 or B3 subfamily, a comparison was carried out of these genes with other known B1 and B3 type hordein genes and the phylogenetic tree was constructed. The results indicated that *Ha2* and *Ha3* were most similar to two B3 type hordein genes respectively and belong to the B3 sub-family. The results provide new information on the variability of B hordein genes.

Key words: Barley, B hordein gene cloning, sequence analysis

INTRODUCTION

The major storage proteins of cereals were called prolamins because of their high proline and glutamine content Prolamins are specifically synthesized and deposited in the endosperm of the developing grain, where they constitute the primary source of nitrogen for the onset of protein synthesis that occurs during subsequent germination and early growth (Wim and Delcour, 2002; Shewry and Halford, 2002; Colot *et al*., 1989). In wheat, the prolamins have been recognized for their important role in determining the nutritive and baking properties of flour and have been classified into two groups, glutenins and gliadins. The prolamins of wheat have been well characterized, with a great deal of different gene sequences deposited in GenBank (http://www.ncbi.nlm.nih.gov). In contrast to wheat, the information about the prolamin genes of barley is little available. The prolamins in barley (*Hordeum vulgare* L.) are usually named as hordeins, which is among the determinants of malting quality, with a negative correlation between malt extract yield and protein content (Molina-Cano *et al*., 2001). They were classified into four main groups: (1) D hordeins, which are homologous to the HMW-GS of wheat, (2) sulphur-poor (S-poor) C hordeins, (3) sulphur-rich (S-rich) B hordeins, which are homologous to the LMW-GS of wheat (Forde *et al*., 1985a; Colot *et al*., 1989) and (4) γ hordeins (Kreis and Shewry, 1989). In barley, the B hordeins are, quantitatively speaking, the major prolamin group, accounting for about 80-90% of the total prolamin fraction (Shewry *et al*., 1995). Genetic studies and extensive characterization at the protein level showed that they are encoded by a multigene family tightly linked at the *Hor2* locus. Further work demonstrated the existence of 10-25 members organized in 2 subfamilies (B1 and B3) (Kreis *et al*., 1983) that comprise at least 85 kb of DNA. Little information is available about the structure, properties and functions of the B hordeins, with only several sequences registered in GenBank.

It is known that Wild barleys, *H. vulgare* ssp. *spontaneum* and *H. vulgare* ssp. *agriocrithon* possess high genetic variation in several useful characters, including earliness, biomass and yield, protein content and a high proportion of resistant genotypes against powdery mildew and leaf rust (Nevo, 2005). Due to its compatible and fully interfertile with the cultivated barley, the wild barley had been used as a source of important genes for cultivar development via interspecific crosses (Dávila *et al*., 1999). In order to learn more gliadins information about wild barleys, B hordeins gene were cloned. In this study, the sequences of three

Corresponding Author: Y.L. Zheng, Triticeae Research Institute, Sichuan Agricultural University, Dujiangyan, 611830, China
Tel: +86 835 2882007 Fax: +86 835 2882336

genes encoding B hordeins from *H. vulgare* ssp. *agriocrithon* and their deduced amino acid sequence were reported. The results will contribute to a better understanding of the genetic variability of the B hordeins family may lead to better exploitation in breeding programs.

MATERIALS AND METHODS

Plant material: The accession ZYM0834 of *H. vulgare* ssp. *agriocrithon* (2n = 2x = 14, 2) was originated from Xizang, China and kindly supplied by Prof. D.Q. Ma (Institute of Crop Germplasm Resources, CAAS) and planted in 2006.

DNA isolated and PCR amplification: CTAB method was adopted for the extraction of genomic DNA from the young leafs of ZYM0834 according to Yan *et al.* (2002). PCR amplification of the ORFs was conducted in 50 μL volume including 150 ng of genomic DNA, 100 mM of each dNTPs, 150 ng of each oligonucleotide primers, 1.5 U ExTaq™ polymerase (TaKaRa), 2 mmol $MgCl_2$ and 5 μL 10×PCR buffer. The pair of primers (P1, P2) were designed based on conserved nucleotide sequences at the 5' and 3' ends of the coding regions of known B hordein genes pBHR184 (GenBank No. X03103) (Forde *et al.*, 1985a) and *pcr31* (GenBank No. X53690) (Vicente-Carbajosa *et al.*, 1992). P1 (5'-ATGAAGAC CTTCCTCATCTTTG-3') contained the start codon (underlined nucleotides). P2 (5'-TTTCTTATCATT AGACACCAAC-3') possessed the tandem stop codons (underlined nucleotides). The cycling parameters were adopted for the PCR reaction according to the following protocol: 94°C for 4 min to pre-denature, followed by 35 cycles of 94°C for 1 min, 60°C for 45 sec and 68°C for 1 min, with a final extension at 68°C for 5 min.

Cloning of PCR products and nucleotide sequence analysis: PCR products were separated in 1% agarose gel (6 V cm^{-1} for 1.5 h in TAE buffer). The desired DNA fragments were purified using a gel extraction kit (Huashun, Shanghai, China), followed by ligation to the pBluescript SK (+/-)-T vector (Stratagene) according to Bielefeldt-Ohmann *et al.* (1997) and then transformed into *E. coli* strain DH10B competent cells. PCR method was used to identify the positive clones. The selected clones were sequenced by a commercial company *TaKaRa* (Dalian, China). The ORFs were translated into amino acid sequences using ORF Finder program in the NCBI (http://www.ncbi.nlm.nih.gov). The sequence analyses were carried out under the program Clustal W 1.81 (Thompson *et al.*, 1994). The MEGA program, version 2, was used to construct the polygenetic tree.

RESULTS

Using the primer pair P1/P2 and genomic PCR amplification, Three DNA fragments, which lengths were about 900, 950 and 980 bp, respectively, were obtained.

As all other prolamin genes, the three novel B hordein genes did not have introns in their coding region and had two typical stop codons at the end (Table 1). Their coding regions contain 900, 945 and 975 bp in nucleotides and can encode three mature proteins with 297, 312 and 322 amino acid residues, respectively. Figure 1 reveals the molecular primary structure of deduced polypeptides from three novel B hordein genes of ZYM 0834. In contrast to the homologous LMW-GS genes of wheat, the N-terminal sequences of the three B hordein genes were blocked, the signal peptide followed by repetitive domain and C-terminal domain.

By the comparison of the amino acid sequences of the three B hordein genes, all genes encoded a predicted, relatively conserved signal peptide with 19 amino acid residues: MKTFLIFALLAIAATSTIA, MKTFLIFALL VIAATSTIA and MKTFLIFALLAIVATSTIA for *Ha1*, *Ha2* and *Ha3*, respectively. Two residue substitutions occurred on two sites (underlined residues in sequence). Immediately following the putative signal peptide there is a central repetitive domain was AC-rich and CC-rich and coded for a large number of glutamine and

Table 1: Characters of the three novel B hordein genes

Gene	Length (bp)	No. of amino acid	pI	Molecular weight (kD)	No. of cys
Ha1	900	297	7.04	33.750	7
Ha2	945	312	8.40	35.299	7
Ha3	975	322	8.30	36.812	8

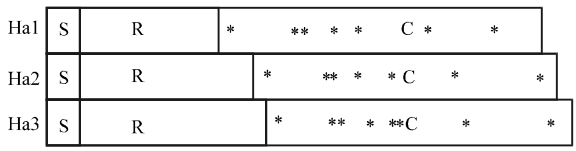

Fig. 1: Illustration of primary structures of three polypeptides deduced from the B hordein genes of ZYM0834. S, R and C indicate signal peptide, repetitive domain and C-terminal domain, respectively. The asterisks refer to the sites of Cys residues

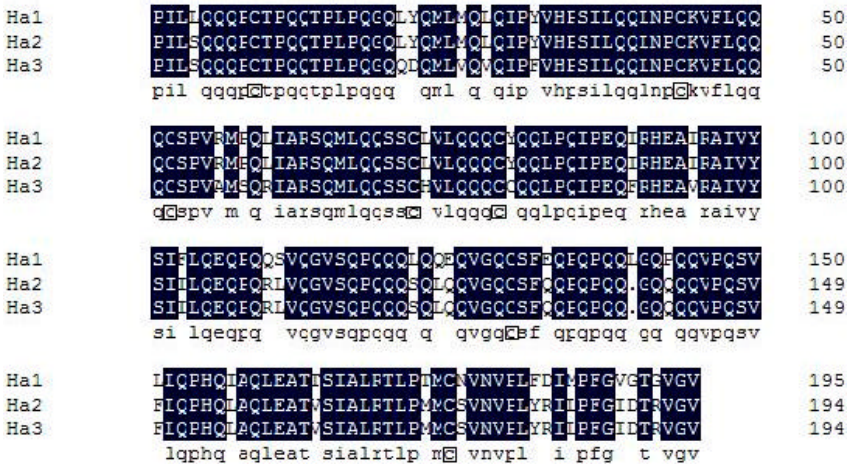

Fig. 2: C-terminal amino acid sequences of the three B hordein genes from ZYM0834. The consensus sequences were blackened and were noted below the alignment. The conservative cysteine residues had been boxed

Table 2: The repeat units of repetitive domain in the three B hordein genes from ZYM0834

Ha1	*Ha2*	*Ha3*
QQPFPQQ	QQPFPQQ	QQPYPQQ
PIPQQ	PFPQQ	PQPFPQQ
PQPYPQQ	PQPYPQQ	PIPQQ
PQQPLPPQQ	PQPFPQQ	PQPYPQQ
PFPQQ	PIPQQPQ	PQPFPQQ
TIPQQPQ	PYPQQ	PIPQQ
PYPHQPQ	PFQPQQ	PQPYPQQ
PYPQQ	PFPQQ	PQPFSQQ
PFPPQQ	TIPQQ	PIPQQ
PFPQQ	PQPYPQQ	PQPYPQQ
PPFWQQQ	PQPYPQQ	PQPFPQQ
PFPQQ	PFPPQQ	PIPQQ
PPFGLQQ	EFPQQ	PQPYPQQ
	PPFWPQQ	PQPFPQQ
	PFPQQ	PFPSQQ
	PPFGLQQ	PFPQQ
		PPFWQQQ

proline residues (Table 2). *Ha1*-encoded protein contains 83 residues, which are 34 glutamine, 31 proline and 7% phenylalanine, *Ha2*-encoded protein contains 99 residues, which are 41% glutamine, 37% proline and 10% phenylalamine and *Ha3*-encoded protein contains 109 residues, which are 43 glutamine, 39 proline and 7% phenylalanine. The size of the central repetitive domain was variable, depending upon the number units of the repeats with 13, 16 and 17 for *Ha1*, *Ha2* and *Ha3*, respectively. The amino acid compositions of repeat units also appeared significant variations among these proteins. Some repeat units, such as PFPQQ, PIPQQ and PYPQQ, were relatively common, while all the repeat units were not conserved in respect of their sequence and length within a gene. No cysteine residue was been found in this domain. The repetitive domain was followed by the C-terminal domain, which was more than half of the protein (Fig. 2). This region was distinguished from repetitive domain by being relatively proline-poor, S-rich and non-repetitive. In this part, all three genes encoded proteins contain 26 glutamine, 11 proline and 3% methionine. *Ha1* and *Ha2* had seven cysteine residues, while *Ha3* had eight cysteine residues; seven of them were completely conservative with the cysteine residues of *Ha1* and *Ha2*. Amino acid sequence comparison indicated that these three genes had a high level similarity (94.02%) and there were a little difference involving single base changes with the consensus sequences.

Table 3: Comparison of *Ha1*, *Ha2* and *Ha3* with those previously characterized B1 or B3 type hordein genes

B hordein	Gene	*Ha1*	*Ha2*	*Ha3*	References
B1 type	*L47.1*	83.43	77.68	72.16	Brandt *et al.* (1985)
	pBHR184	89.88	85.55	83.01	Forde *et al.* (1985a)
	pcr47	89.88	87.13	76.97	Vicente-Carbajosa *et al.* (1992)
B3 type	*pcr31*	89.89	89.77	74.85	Vicente-Carbajosa *et al.* (1992)
	pB7	74.08	69.83	74.74	Forde *et al.* (1985b)

The molecular weights and the pI values of the polypeptides encoded by the three genes were calculated by the mean molecular weight of individual amino acids and the program Clustal W 1.81 (Thompson *et al.*, 1994). The predicted molecular weights of the polypeptides encoded by *Ha1*, *Ha2* and *Ha3* were 33.750, 35.299 and 36.812 kDa, respectively. The pI values of the polypeptides encoded by *Ha1*, *Ha2* and *Ha3* were 7.04, 8.40 and 8.30, respectively.

Optimal alignment of nucleotides of the three novel B hordein genes derived from ZYM0834 were compared between each other and to those of previously characterized genes *pBHR184*, *pcr47*, *L47.1* (B1 type) and *pB7*, *pcr31* (B3 type) (Table 3). *Ha1* had relatively

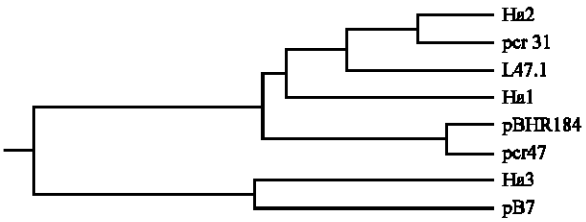

Fig. 3: Polygenetic tree based on alignment of the DNA sequences of the representative B hordein genes

higher similarity (89.88 and 89.89%) both with B1 type (*pBHR184*, *pcr47*) and B3 type (*pcr31*). *Ha2* had relatively higher similarity (89.77%) with B3 type (*pcr31*). While *Ha1* had relatively higher similarity (83.01%) with B1 type (*pBHR184*). The classification of *Ha1*, *Ha2* and *Ha3* in relation to B1 or B3 type hordein genes was carried out through an analysis of the their ORFs (Fig. 3). The *Ha2* and *pcr31* (B3 type), The *Ha3* and pB7 (B3 type) were most similar and were classified into unique branch, respectively. The results suggested that *Ha2* and *Ha3* may be more closely related to the B3 type, while there was no significant evidence to classified *Ha1* into B1 or B3 type hordeins.

DISCUSSION

It is the hordeins, the major fraction of the endosperm storage proteins that are mainly responsible for the negative correlation between malt extract yield and barley protein content (Jun-Cang Qi *et al*., 2005; Roxana *et al*., 2006). The hordeins are soluble in alcohol /water mixtures and, on average, account for up to 60% of the total grain nitrogen. They are classified into 4 heterogenous groups, or families of polypeptides and are encoded by the genes: *Hor2* (B-fraction), *Hor1* (C-fraction), *Hor3* (D-fraction) and *Hor5* (γ-fraction), located on barley chromosome 5 (homoeologous chromosome 1H). The B and C hordein groups together account for about 95% of the total hordein fraction (Henriette *et al*., 1983; Molina-Cano *et al*., 2001).

Three different genomic clones, *Ha1*, *Ha2* and *Ha3*, from *H. vulgare* ssp. *agriocrithon* were isolated and sequenced. Sequence analysis of these clones indicated that their coding regions, as other B hordein genes, were not interrupted by introns. The open reading frames encoded three proteins with 297 residues (Mr 33.808), 312 residues (Mr 35.913) and 322 residues (Mr 36.873) for *Ha1*, *Ha2* and *Ha3*, respectively. The amino terminal region of the protein has many of the characteristics of a signal peptide, including a charged residue (isoleucine) near the N-terminus and a core of hydrophobic residues (FLIFALLA/VIA/VA, A/V designate the two substitutions). The presence of a signal peptide would be consistent with the evidence that the B hordeins are synthesized on the rough endoplasmic reticulum and deposited in protein bodies (Forde *et al*., 1985a). However, because the N-terminus of the B hordeins is blocked, (Shewry *et al*., 1980), the short N-terminal domain in other S-rich prolamins is absent from these three B hordein polypeptides. The N-terminal sequence of the mature protein is not known and we can not assign the site of signal peptide cleavage with certainty. Nevertheless, comparison with the mature sequences of the homologous storage proteins from wheat suggested that the cleavage would occur between residues 19 and 20 (Forde *et al*., 1985a). Consequently, the three mature proteins would therefore consist of 278 residues (Mr 31.829), 293 residues (Mr 33.907) and 303 residues (Mr 34.866) for *Ha1*, *Ha2* and *Ha3*, respectively, which agree well with estimates based on direct analysis of the B hordeins (Shewry *et al*., 1980). Immediately following the putative signal peptide there is a region of the protein that is extremely rich in praline and glutamine. This part of the protein was previously identified as domain 1, one of two domains of primary structure in B hordein polypeptides, which is also characterized by a series of tandem repeats. The size of this domain was variable, depending upon the number units of the repeats. The amino acid compositions of repeat units also appeared significant variations among these proteins. Some repeat units, such as PFPQQ, PIPQQ and PYPQQ, were relatively common. The domain 1 was followed by the domain 2, which distinguished from domain 1 by relatively proline-poor, S-rich and non-repetitive. All of the cysteine residues (seven for *Ha1* and *Ha2*, eight for *Ha3*) were presence in domain 2.

The initial classification of B hordeins into B1 and B3 types was based on protein characteristics including Mr, pI, CNBr mapping patterns and organization of the repetitive domain (Shewry and Tatham, 1990). Nevertheless, the limited information on nucleotide sequences, in particular the lack of complete B3 type coding sequences, make it difficult to classify any new gene simply by sequence comparisons (Vicente-Carbajosa *et al*., 1992). The results presented above provide new information on the variability of B hordein genes, in particular, the information about B3 type hordeins.

ACKNOWLEDGMENTS

The wild barley accessions were kindly supplied by Prof. D.Q. Ma (Institute of Crop Germplasm Resources, CAAS). This work was supported by the Hi-Tech

Research and Development (863) Program of China (2003AA207100) and the Foundation for the Author of National Excellent Doctoral Dissertation of PR China (200357). Y.L. Zheng was supported by Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT), China.

REFERENCES

Bielefeldt-Ohmann, H. and D.R. Fitzpatrick, 1997. High efficiency T-vector cloning of PCR products by forced a tagging and post-ligation restriction enzyme digestion. Biotechniques, 60: 822-822.

Brandt, A., A. Montenbault, V. Cameron-Mills and S.K. Rasmussen, 1985. Primary structure of a B1 *hordein* gene from barley. Carlsberg Res. Commun., 50: 333-345.

Colot, V., D. Bartels, R. Thompson and R. Flavell, 1989. Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. Mol. Gen. Genet., 216: 81-90.

Dávila, J.A., Y. Loarce, L. Ramsay, R. Waugh and E. Ferrer, 1999. Comparison of RAMP and SSR markers for the study of wild barley genetic diversity. Hereditas, 131: 5-13.

Forde, B.G., A. Heyworth, J. Pywell and M. Kreis, 1985a. Nucleotide sequence of a b1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucleic Acids Res., 13: 7327-7339.

Forde, B.G., M. Kreis, M.S. Williamson, R.P. Fry and J. Pywell *et al*., 1985b. Short tandem repeats shared by B- and C-hordein CDNAS suggested a common evolutionary origin for two groups of cereal storage protein genes. EMBO J., 4: 9-15.

Henriette, G., A. Bente and D. Hans, 1983. Synthesis of the major storage protein, hordein, in barley. Planta, 159: 60-65.

Jun-Cang, Q., C. Jin-Xin, W. Jun-Mei, W. Fei-Bo, C. Lian-Pu and Z. Guo-Ping, 2005. Protein and hordein fraction content in barley seeds as affected by sowing date and their relations to malting quality. J. Zhejiang Univ. Sci. B, 6: 1069-1075.

Kreis, M., S. Rahman, B.G. Forde, P.R. Pywell, P.R. Shewry and B.J. Miflin, 1983. Sub-families of hordein MRNA encoded at the *Hor2* locus of barley. Mol Gen. Genet., 191: 194-200.

Kreis, M. and P.R. Shewry, 1989. Unusual features of seed protein structure and evolution. Biol. Essays, 10: 201-207.

Molina-Cano, J., J.P.L. Polo, E. Romera, J.L.A. raus, J. Zarco and J.S.S. Wanston, 2001. Relationships between barley hordeins and malting quality in a mutant of cV. Triumph Genotype by Environ. Interaction of hordein content. J. Cereal Sci., 34: 285-294.

Nevo, E., 2005. Population Genetic Structure of Wild Barley and Wheat in the Near East Fertile Crescent: Regional and Local Adaptive Patterns. Springer, Netherlands, ISBN: 978-1-4020-2358-3,.

Roxana, S., P. Pablo and L.A. Jose, 2006. Hordein composition as affected by post-anthesis source-sink ratio under different nitrogen availabilities. J. Cereal Sci., 44: 113-116.

Shewry, P.R., J.M. Field, M.A. Kirkman, A.J. Faulks and B.J. Miflin, 1980. The extraction, solubility and preliminary characterization of two groups of barley storage protein (hordein) polypeptides. J. Exp. Bot., 31: 393-407.

Shewry, P.R. and A.S. Tatham, 1990. The prolamin storage protein of cereal seeds: Structure and evolution. Biochem. J., 267: 1-12.

Shewry, P.R., J.A. Napier and A.S. Tatham, 1995. Seed storage protein: Structures and biosynthesis. Plant Cell, 7: 945-956.

Shewry, P.R. and N.G. Halford, 2002. Cereal seed storage proteins: Structures, properties and role in grain utilization. J. Exp. Bot., 53: 947-958.

Thompson, J.D., D.G. Higgins and T.J. Gibson, 1994. Clustal w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22: 4673-4680.

Vicente-Carbajosa, J., D.R. Beritashvili, A.S. Kraev and K.G. Skryabin, 1992. Conserved structure and organization of B hordein genes in the *Hor2* locus of barley. Plant Mol. Biol., 18: 453-458.

Wim, S.V. and J.A. Delcour, 2002. Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality. Critical Rev. Food Sci. Nutr., 42: 179-208.

Yan, Z.H., Y.F. Wan, K.F. Liu, Y.L. Zheng and D.W. Wang, 2002. Identification of a novel HMW glutenin subunit and comparison of its amino acid sequence with those homologous subunits. Chinese Sci. Bull., 47: 220-225.